IJSRCSEIT International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/IJSRCSEIT

Diagnosing Cancer using Data Mining: A Critical Review of Techniques

Qamar Rayees Khan
Department of Computer Sciences, BGSB University, Rajouri, Jammu & Kashmir, India

ARTICLEINFO

Article History:

Accepted: 01 Nov 2016
Published: 11 Dec 2016

Publication Issue
Volume 1, Issue 3
November-December-2016
Page Number
94-103

ABSTRACT

Computer science has revolutionized the whole world by its various application areas. Data mining is one of the sub areas of Computer Science, which extracts hidden information from the data. Data mining is used for diagnosing various diseases. In this paper, a critical review is performed on the various techniques that are used for diagnosing cancer. Various preprocessing techniques are also discussed in this work. Median, Mean and Wiener filtering techniques are also provided as a base for this work.

Keywords— Data Mining, Analytics, Diseases, Cancer, Wiener filtering.

I. INTRODUCTION

Data Mining (DM) is a convenient way of extracting patterns which represents knowledge implicitly stored in large datasets, and focuses on issues relating to their feasibility, usefulness, effectiveness and scalability. It can be viewed as an essential step in the process of knowledge discovery. There are different stages for data preprocessing, namely, (i) Data cleaning (iii) Data integration (iv) Data selection (v) Data transformation, after which they are made ready for the mining task. DM is the process of discovering meaningful new correlation, patterns and trends by sifting through large amounts of data, using pattern recognition, as well as, statistical and mathematical techniques. Started as little more than a dry extension of DM techniques, it is now attributed with bringing important contributions in the crucial fields of investigations and among traditional sciences such as Astronomy, High Energy Physics, Biology and Medicine. It has always provided a rich source of applications, developed on the base of DM [1,2,3].
DM is a promising discipline and has broad applicability. It can be applied in various areas, merging multiple intertwined disciplines, including statistic, machine learning, shape identification, database organization, information retrieval, World Wide Web, visualization, and many other application domains [4,5,6]. It has produced large progress in the past two decades. Evolution of DM applications and its implications are manifested in the areas of data management in healthcare administrations, epidemiology, patient care and intensive care systems, significant image analysis for information extraction and automatic identification of unknown subjects. DM is applied to find useful

patterns, to help with the essential tasks of medical scrutiny and treatment. DM techniques are used to find new, hidden and useful patterns of knowledge from databases. The numerous mining functions are association rules, classification, prediction, and clustering. In DM, it is well known that classification techniques are more suitable to larger databases. Therefore, classification techniques are used on a larger mammographic database to extract more features from the images, thereby, predicting breast cancer. In particular, a classification model based on association rules becomes more accurate with a larger dataset of images. In addition, more features attached to the images such as age, with/without children and so on, could be interesting and relevant as additional attributes for classification. In the case of Association Rule Mining approach, image split in more windows could improve the detection by localization of the tumor. Data analysis systems, intended to assist a physician, are highly desirable to be accurate, human interpretable and balanced, with a degree of confidence associated with final decision. In cancer prognosis, such systems estimate recurrence of disease and predict survival of patient, hence resulting in improved patient management. To develop such a prognostic system, fuzzy decision trees are used.

DM techniques and Artificial Neural Networks (ANNs) are considered as effective methods for diagnosis of breast cancer. A variety of different parallelization strategies have been considered for ANNs. Some of such strategies are Exemplar Parallelism (EP), Block parallelism (BP), Neuron Parallelism (NP). Experiments are conducted by considering the single and Multi-layer Neural Network models. Back propagation algorithm with momentum and variable learning rate is used to train networks and Memory Level Parallelism (MLP) is adopted [7]. Then a modified Normalized Cut approach considering the weighted neighborhood gray values is proposed, to partition the ROI into clusters and get the initial boundary. In addition, a regional fitting active contour model is used to adjust the few inaccurate initial boundaries for the final segmentation. Usually, three textures and five morphologic features are extracted from each breast tumor; whereby a highly efficient affinity propagation clustering is used tofulfill the malignancy and benign classification for an existing database without any training process. This system needs no training procedure or manual interference; however, performs best in detection and classification of ultrasonic breast tumors, while having the lowest computation complexity [8]. This paper consists of five sections, Section one gives an overview of data mining and its importance. In section two related work regarding cancer diagnosing is provided. In section three various cancer diagnosing techniques are discussed, section four provides the preprocessing techniques and section five concludes the critical review.

II. LITERATURE REVIEW

Nowadays, a large amount of medical data, storing patients' medical history, is collected during health care. The analysis of these medical data collected, is a challenging task for health care systems, since vast amount of interesting knowledge can be automatically mined to effectively support both physicians and health care organizations [9]. There are numbers of research works carried out by different people to find the medical data. Presently, there are a number of techniques used and applied to analyze medical diseases. Visalatchi, G, et al., in [10] stated there are different DM classification techniques, which can be used for the identification and prevention of diabetes disease among patients. This article describes some classification techniques in DM to predict diabetes disease in patients, namely C4.5, Support Vector Machine (SVM), k-Nearest Neighbour (k-NN), Naive Bayes, and Apriori algorithm. These techniques are compared with each other for their accuracy, on the basis of different attributes of diabetes disease. One of the algorithms has accuracy above 85%, that is, C4.5 algorithm. C4.5 algorithm was the best amongst the five techniques. Aqueel Ahmed et al., in [11] stated that Decision Tree algorithms and SVM perform classification are more accurate than the other methods. They reported that Decision Tree and SVM perform classification more accurately than the other methods and were able to achieve 92.1 % and 91.0 % accuracy. Venkatesan P. et al., in [12] discussed and compared three Decision Tree algorithms. They classified tuberculosis patients' responses under randomized clinical trial condition, which was carried out by them. Classification of patients' responses to treatment is based on bacteriological and radiological methods. Three Decision Tree approaches, namely C4.5, Classification and Regression Trees (CART), and Iterative dichotomizer3 (ID3) methods were used for classification of patients' responses. The result shows that C4.5 Decision Tree algorithm performs better than CART and ID3 methods. Kirthika et al., in [13] discussed about the heart disease database, which is preprocessed to make the mining process more efficient. The preprocessed data are clustered using the k-Means clustering algorithm for clustering relevant data in the database. Maximal Frequent Item-set Algorithm (MAFIA) is used for mining maximal frequent patterns in the heart disease database. The frequent patterns can be classified using C4.5 algorithm, as the training algorithm uses the concept of information entropy. They concluded in their results that the designed prediction system is capable of predicting heart attack accurately. Classification of diabetes disease using SVM algorithm is done by Anuja kumar et al. in their proposed work [14]. SVM with Radial basis function Kernel is used for classification. Performance parameters such as the classification accuracy, sensitivity, and specificity of SVM and RBF have found to be high, thus making it a good option for the classification process. In future, the performance of SVM classifier can be improved by feature subset selection process. Vanaja S. and Ramesh Kumar in [15] discussed about data constraints such as volume and dimensionality problems. This paper also discusses the new features of C5.0 classification algorithm over C4.5 and performance of the classification algorithm on high dimensional datasets. In this analysis, C5.0 algorithm is applied to the high dimensional dataset and it must incorporate any one of the best feature selection algorithm for better performance.

Velmurugan, T and Dharmarajan, A. in [16] identified the applications of innovative and special approaches of clustering algorithms, principally, for the medical domain. They conclude that from the various applications suggested by several researchers the performance of k-Means algorithm is well suited for this type of medical dataset analysis. Most of the researchers are using the k-Means algorithm; also it is more suitable than other algorithms in the medical data set. Kalyani P., in [17] explored and analyzed the problem of partitioning medical data. They enhanced the existing traditional algorithms (k-Means, Density-Based Spatial Clustering Of Applications with Noise (DBSCAN) and FCM) and proposed k-Means, DBSCAN and FCM Clustering Algorithm performance evaluation. Raikwal J. and Saxena K. in [18] did a research over a medical data set and they made a comparison between k-NN and SVM results, and after implementing the two algorithms, they showed that k-NN is quite a good classifier when applied over small data set; however, its accuracy decrease when it is applied over a large data set and show poor results. Karegowda A. G., Jayaram M., and Manjunath A., [19] wrote a paper on using cascading k- Means clustering and k-NN classifier on diabetic patient and the result was quite good. Due to its simplicity and high accuracy, k-NN approach has been used in different data analysis applications such as pattern recognition, DM, databases and machine learning. Sujatha N. and Iyakutty K. in [20] discussed the significant role of k-NN algorithm in the field of medical diagnosis and it was used for uterus cancer diagnosis. The classification work carried out by researchers was done with k-Means clustering algorithm for uterus cancer dataset. The experimental results demonstrated 21 that their proposed work is very effective in producing desired clusters of the given dataset, as well as, diagnosis. These algorithms are very much useful for image classification, as well as, extraction of objects in the medical domain. Senguttuvan A. et al., in [21], represented computational complexity of binary dataset under five different clustering algorithms, namely k-Means, C-Means, Mountain Clustering, Subtractive Method, and Extended Shadow Clustering Algorithms. The researchers implemented and tested them against a medical problem of heart disease diagnosis. The conclusion in their research work shows that the performance of k-Means is good, when implemented. k-NN approach has been used in different data analysis applications such as pattern recognition, DM, databases and machine learning due to its simplicity and high accuracy. It has been recognized as one of the top 10 algorithms in DM [22]. The Pima Indian diabetes dataset is complex due to its missing values. A class wise, k-NN algorithm have been designed and tested against the Pima Indian diabetes dataset. Here testing data are classified into class label corresponding to the lowest distance. The accuracy achieved for k-NN is 78.16% [23]. Arpana M. A. and Prathiba Kiran [24] presented that, currently, digital mammography is the most efficient and widely used technology for early breast cancer detection. The major diagnosing elements such as masses and lesions in the digital mammograms are noisy and of low contrast. The aim of the proposal was to enhance the mammogram images by reducing the noise using Median Filter, Image Sharpening and Image Smoothing. The data clustering algorithm i.e. FCM clustering is used to segment the ROI from which various statistical, gradient and geometrical features are extracted. The features extracted from a few images of the database are used to train the neural networks for classification. The evaluated algorithm is tested on the digital mammograms from the Mammogram Image Analysis Society (MIAS) database. Their experimental results show that the breast region extracted by the presented algorithm approximately follows that extracted by an expert radiologist. The detected mass is classified as normal or abnormal. Further, abnormal can be classified into benign or malignant.

Table 1: Related Work

Author	**Year**	**Contribution in this Direction**	**Research Gap**
Sujatha et al.	2010	Performed computational complexity of binary dataset under five different clustering algorithms, namely k- Means, C-Means, Mountain Clustering, Subtractive Method, and Extended Shadow Clustering Algorithms.	More algorithms can be used.
Raikwal et al.	2012	Made a comparison between k-NN and SVM results, and after implementing the two algorithms, they showed that k-NN is quite a good classifier when applied over small data set.	Small dataset have been used for experimentation.
Arpana et al.	2014	Showed that digital mammography is the most efficient and widely used technology for early breast cancer detection. Their experimental results showed that the breast region extracted by the presented algorithm approximately follows that extracted by an expert radiologist.	Multi Class classification can be done in future.
Velmurugan et al.	2015	Identified the applications of innovative and special approaches of clustering algorithms	Other Algorithms can be used in place of K-means

III.CANCER DIAGNOSING TECHNIQUES

Different kinds of cancer screening methods have been found to detect cancer. Some of these tests are Mammography, Ultrasound, Magnetic Resonance Imaging (MRI), and Computed Tomography (CT) Mammogram. Ultrasound and X-ray mammogram are the most widely used techniques. This approach uses only mammography method to discover tumor in mammogram images [25].

3.1. Mammography Technique

Digital Mammography is also known as fully upgraded by Full Field Digital Mammography (FFDM). Here, X-ray film is replaced by a solid state detector. Solid state detectors convert x-rays' film image into mammographic pictures (into electrical signals.) of the breast. Signals are used to capture the breast's inner part, so as to produce a special digital image. In Mammography test, only one picture can be taken at a time and also only one side of the breast will be captured. Compression of breast causes overlapping of tissues and hides certain details. Therefore, Mammography sometimes does not show malignant tumor. The main disadvantage is compression of the breast. It may be a painful procedure such as CAD. CAD systems search digitized mammographic images for abnormal areas by density, mass or calcification that may indicate the presence of cancer. The CAD system highlights these areas on the images, alerting the radiologist to carefully assess this area. CAD produces digitally acquired mammogram. CAD is very helpful to radiologist to detect cancer. Breast Tomo synthesis is a 3 dimensional picture representation of breast using X-rays. It is not considered as standard testing of breast cancer. The main drawback in this testing is that the device is not available in many hospitals. Breast Tomosynthesis overcomes the above said disadvantage of Mammography. Breast Tomosynthesis takes multiple images of breast at many different angles. Breast Tomosynthesis has an X-ray tube arc. During the test, X-ray tube arc is placed around the breast and it takes 11 highly clear 3 dimensional (3D) images.

3.2. Ultrasound Technique

Ultrasound is one of the techniques used to detect breast cancer. Ultrasound is otherwise called as Sonography or Ultrasound Screening. The Ultrasound device consists of computer, electronic transducer, ultrasound gel and video display. The 11 transducer is used to scan the body. The transducer is a handheld device and it has attached microphone. Ultrasound test is non-invasive test and painless. Ultrasound gel exposes the body to high frequency sound waves. During Ultrasound testing, it does not emit any iron radiation as in Mammography. The ultrasound gel is applied in the suspicious breast region. Transducer helps display of the internal structure of the breast and movements of an internal organ, showing the flow of blood vessels. Ultrasound is less expensive. Doppler Ultrasound tests the blood flow through blood vessels and also shows the body's major arteries, veins, abdomen, arms and neck. During breast Ultrasound test, the sonographer knows the blood flow in the breast. Therefore, they use Doppler Ultrasound test and also find the lack of flow in breast mass. The transducer produces some sounds like frequency sound waves and sends into the body to test.

3.3. Magnetic Resonance Imaging (MRI)

MRI is full of magnetic field and radio frequency pulses. MRI produces strong magnetic rays into the body [26,27]. It is less expensive. Computer MRI tests the internal organs, soft tissues, bones and internal structure. Images are shown on computer screen and transmitted into electronic signals. Then the details are printed or copied in a Compact Disc (CD) in image format. MRI does not produce any iron radiation. MRI test helps to find how large the cancer is and where the suspected muscles are underlying. MRI is capable of capturing the images of both the breasts simultaneously. Any abnormality tumor or lymph node in the armpit is detected easily. MRI finds current stage of cancer and also abnormalities. MRI easily detects dense breast tissue in younger women. Positron Emission Tomography (PET) helps to detect cancer area and the body's cells first. PET scans start with an injection of radio pharmaceutical called Flu Deoxy Glucose (FDG). During the PET scan, gamma rays are emitted by FDG. FDG are recorded by the PET scanner and images are reconstructed and reviewed. This helps to identify the suspected malignancy. If a physician can find suspicious area, it will accumulate the signals stronger in suspected tissues [28]. MRI Artifacts: Both patient and technical factors may lead to unwanted artifacts in breast MRI. The use of a properly functioning high-field-strength MRI system and an 12 optimal imaging protocol are important for avoiding artifacts. Difficulties involving breast positioning and selection of an appropriate imaging volume can be overcome by training MRI technologists and providing them with imaging-based feedback. Other important steps include using optimal fat suppression and minimizing patient motion. Recognizing image quality problems and imaging artifacts (for example, motion, suboptimal fat suppression, metallic susceptibility, chemical shift, image wrap, and RF (Radio Frequency) noise) are essential for maintaining optimal breast imaging. Once recognized, many artifacts can be corrected by the MRI technologist and radiologist working together, to optimize imaging techniques. In this work, Mammography technique is used for detecting breast cancer. Mammogram is a basic test to detect breast cancer. During the Mammogram test, the iron radiation that goes into the breast shows internal parts of the body and also the suspicious region.

IV.PRE-PROCESSING TECHNIQUE

Data preprocessing is one of the most critical steps in DM process which deals with the preparation and transformation of the initial dataset. Data preprocessing is used to remove the unwanted data and information available in the data set. It is a process of DM which deals with training and transformation of the original data set. The phrase "Garbage In, Garbage Out" (GIGO) is particularly pertinent to DM and machine learning approaches to remove irrelevant information from the data set [29].Image preprocessing through noise is characterized by its pattern and probabilistic characteristics. There are different types and kinds of noises. Some of them are Gaussian noise, Impulse noise (Salt and Pepper noise), Poison noise, Uniform Noise, and Speckle noise.

a) **Gaussian noise** is a statistical noise that has its probability density function equal to that of the normal distribution, which is also known as the Gaussian distribution. In other words, the values that the noise can take on being Gaussian distributed. A special case is White Gaussian noise, in which, the values at any pair of times are identically distributed and statistically independent and hence, uncorrelated. In applications, Gaussian noise is most commonly used as additive white noise to yield additive white Gaussian noise [30,31].

b) **Impulse Noise (Salt & Pepper Noise):** It is also known as Impulse Noise. Black and white dots appear in the image as a result of this noise and hence, the name Salt and Pepper noise. When only black spots are present it is called Pepper noise and if only white spots are present then it is called Salt noise. The reason behind this type of a noise is due to sharp and sudden changes of image signal. An effective method of noise reduction for this type of noise is a Median Filter or a Morphological Filter. For reducing either Salt Noise or Pepper noise, but not both, a contra harmonic mean filter can be effective.

c) **Uniform Noise:** It is also called Quantization noise. It is caused by quantizing the pixels to a number of discrete levels called as Quantization noise. Its distribution is approximately uniform distribution. It is signal dependent, but, can be signal independent if other noise sources are big enough to cause dithering (Dither is an intentionally applied form of noise used to randomize quantization error, preventing large-scale patterns such as color banding in images) or applied explicitly. This type of noise provides the most neutral or unbiased noise.

d) **Poisson Noise:** Poisson or Shot Photon noise is the noise that can cause, when the number of photons sensed by the sensor is not sufficient to provide detectable statistical information. This noise has a root mean square value proportional to square root intensity of the image. Different pixels suffer independent noise values. On practical grounds photon noise and other sensor based noise corrupt the signal in different proportions.

e) **Periodic noise:** Periodic noise is usually caused by interference between electronic components. It has a fixed frequency, phase and amplitude. It typically arises from interference during image acquisition. Periodic noises have spurious signals that create repetitive patterns on images.

f) **Speckle Noise:** Also called as Multiplicative noise, it can be modeled by random value multiplications with pixel values of the image expressed as, J = I + n*I J = speckle noise distribution image, I = input image, n = uniform noise image with mean o and variance v. This noise deteriorates the quality of active radar and Synthetic Aperture Radar (SAR) images. Various methods are there for Speckle noise reduction and enhancement.

Data preprocessing methods are divided into Data Cleaning, Data Integration, Data Transformation and Data Reduction. Image pre-processing techniques are necessary to improve the image quality, to make it ready for further processing by taking away or reducing unrelated and surplus parts in the background of the medical images that is difficult to understand. The common characteristics of the medical images such as unknown noise, poor image contrast, inhomogeneity, weak boundaries and unrelated parts will affect the content of medical images. This problem is rectified by preprocessing techniques. Preprocessing done in medical images produce better image quality for segmentation and feature extractions. Preprocessing steps deals with image enhancement, noise and special mark removal in problem retrieved using filtering techniques. The noise and high frequency components are removed by filters. Preprocessing image filtering techniques are Mean Filter or Average Filter, Median Filter, Adaptive Median Filter, Wiener Filter and Gaussian Filter [32].

4.1. Median Filter

Median filter is efficient in removing Salt and Pepper noise. Median tends to keep the sharpness of image edges while removing the noise. Median Filter is a nonlinear digital filtering technique which is used for noise reduction in image. It is an algorithm that is useful for the removal of impulse noise [33]. There are several variations of median filter such as Centre-Weighted Median Filter, Weighted Median Filter, Max-Median Filter. The effect of the size of the window increases when Median Filter remove noises effectively.

4.2. Mean Filter

Mean Filter replaces each pixel with the average value of intensities in its neighborhood. It can locally reduce the variance and is easy to implement. It has the effect of smoothing and blurring the image and is optimal for additive Gaussian noise in the sense of mean square error. Speckled image is a multiplicative model with nonGaussian noise and therefore, the simple mean filter is not effective in this case. In order to alleviate the blurring effect, the adaptive Mean Filters have been proposed to achieve a balance between straightforward averaging and all-pass filtering. They adapt the properties of the image locally and selectively remove speckles from different parts of the image. The use of local image statistics such as Mean, Variance and Spatial Correlation is to effectively detect and preserve edges and features. Speckle noise is removed by replacing it with a local mean value. The adaptive mean filters outperform mean filters, and generally reduce speckle while preserving the edges [34].

4.3. Wiener Filtering

Wiener filtering method requires information on the spectra of noise and the original signal, and it works well only if the underlying signal is smooth. Wiener method implements spatial smoothing and its model complexity control correspond to choose the window size. Wiener filtering is able to achieve significant noise removal when the variance of noise is low. They cause blurring and smoothening of the sharp edges of the image. Mammograms are medical images that are complicated to interpret. Hence, preprocessing is essential to improve the quality. In this research work, Median Filter and Gaussian Filter are used for pre-processing. In addition to the filtering techniques used for preprocessing in the ROI, Inverse methods and Boundary Deduction methods are used to get better image quality as a final preprocessing result.

V. CONCLUSION

In this work, a detailed review is provided to diagnose the cancer using image dataset. Techniques like Mammography, Ultrasound and Magnetic Resonance Imaging are discussed in detail. Various preprocessing techniques are also explained in detail. Filtering techniques like Mean, Median and Wiener are also discussed. There is a lot of scope in doing research on various datasets.

VI. REFERENCES

[1]. Eze, Udoka Felista, Adeoye, Olufemi Sunday and Ikemelu, Chinelo Rose-keziah, “Industry Wide Applications of Data Mining”, International Journal of Advanced Studies in Computer Science and Engineering, Vol. 2 (3), pp. 28-32, 2014.

[2]. Naveen A. and Velmurugan T., "A Survey on Medical Images Extraction using Parallel Algorithm in Data Mining", International Conference on Information, System and Convergence Applications, pp. 86-91, 2015.

[3]. Sanpawat Kantabutra and Alva L. Couch, “Parallel k-Means Clustering Algorithm on Nows”, NectecTechnical Journal, Vol. 1 (6), pp. 243-24, 2000.

[4]. Aruna S., Rajagopalanand S. P., and Nandakishore L. V., “Knowledge Based Analysis of Various Statistical Tools in Detecting Breast Cancer”, Computer Science &Information Technology, Vol. 2, pp. 37-45, 2011.

[5]. Bharati and M. Ramageri., “DM Techniques and Applications”, Indian Journal of Computer Science and Engineering”, Vol.1 (4), pp. 301-305, 2010.

[6]. Hiral N. Pokar and Poorvi H. Patel., “Survey on Different Techniques Used For Detection of Malignancy in Mammograms of Breast Cancer”, National Conference on Emerging Trends in Computer, Electrical & Electronics, 2015.

[7]. Burke H. B., Goodman P. H., Rosen D. B., Henson D. E., Weinstein J. N., Harrell F. E., and Bos twick D. G., “Artificial Neural Networks Improve the Accuracy of Cancer Survival Prediction”, CANCER, Vol. 79(4), pp. 857-862, 1997.

[8]. Carona Y. P., Makrisa and Vincent N., “Use of Power Law Models in detecting Region of Interest”, Journal of Pattern Recognition Society Elsevier, Vol. 9 (40), pp. 2521–2529, 2007.

[9]. Xin Xiao and Silvia Chiusano, “Analysis of Medical Treatments Using Data Mining Techniques”, IEEE Intelligent Informatics Bulletin, Vol. 15 (1), pp. 30- 31, 2014.

[10]. Visalatchi G., Gnanasoundhari S. J., Balamurugan M., “A Survey on Data Mining Methods and Techniques for Diabetes Mellitus”, International Journal of Computer Science and Mobile Applications, Vol. 2 (2), pp. 100-105, 2014.

[11].Aqueel Ahmed and Shaikh Abdul Hanna, “Data Mining Techniques to Find out Heart Diseases: An Overview”, International Journal of Innovative Technology and Exploring Engineering, Vol. 1 (4), pp. 18 -23, 2012.

[12].Venkatesan P., Yamuna N. R., “Treatment Response Classification in Randomized Clinical Trials through Decision Tree Approach”, Indian Journal of Science and Technology, Vol. 6 (1), pp. 3912-3917, 2013.

[13].Kirthika B., Malathi P., Yashwanthi Sivakumari C. L., and Sudharsan P., “A Comparative Analysis of De - noising Techniques in Ultrasound B mode Images”, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 3 (1), pp. 5136-5140, 2014.

[14].Anuja Kumari V. and Chitra R., “Classification of Diabetes Disease Using Support Vector Machine“,International Journal of Engineering Research and Applications, Vol. 3 (2), pp. 1797-1801, 2013.

[15].Vanaja S. and Ramesh Kumar K., “Performance Analysis of Classification Algorithms on Medical Diagnoses: A Survey”, Journal of Computer Science, Vol. 1 (11), pp. 30.52, 2015.

[16].Velmurugan T. and Dharmarajan A., “Clustering Lung Cancer Data by k Means and k-Medoids Algorithms”, International Conference on Information and Convergence Technology for Smart Society, pp. 21-24, 2015.

[17].Kalyani P., “Approaches to Partition Medical Data Using Clustering Algorithms“, International Journal of Computer Applications, Vol. 49 (23), pp. 7-10, 2012.

[18].Raikwal J. and Saxena K., “Performance Evaluation of SVM and K-Nearest Neighbor Algorithm over Medical Data set”, International Journal of Computer Applications, Vol. 50, pp. 35-39, 2012.

[19].Karegowda A. G., Jayaram M., and Manjunath A., “Cascading k-means Clustering and k-Nearest Neighbor Classifier for Categorization of Diabetic Patients”, International Journal of Engineering and Advanced Technology, Vol. 1, pp. 147- 151, 2012.

[20].Sujatha N. and Iyakutty K., “Refinement of Web usage Data Clustering from K-means with Genetic Algorithm”, European Journal of Scientific Research, Vol. 50, pp. 478–90, 2010.

[21].Senguttuvan A., Krishna Pramodh D., and Rao Venugopal K., “Performance Analysis of Extended Shadow Clustering Techniques and Binary Data Sets Using K- means Clustering”, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 2 (8), pp. 51–62, 2012.

[22].Ranshul Chaudhary, Prabhdeep Singh and Rajiv Mahajan., “A Survey on Data Mining Techniques”, International Journal of Advanced Research in Computer and Communication Engineering, 3 (1), 2014, pp. 5002-5003.

[23].Angeline Christobel Y. and Sivaprakasam P.,"A New Class-Wise k Nearest Neighbor (CK-NN) Method for the Classification of Diabetes Dataset", International Journal of Engineering and Advanced Technology, Vol. 2 (3), pp. 396-400, 2013.

[24].Arpana M. A. and Prathiba Kiran., “Feature Extraction Values for Digital Mammograms”, International Journal of Soft Computing and Engineering, Vol. 4 (2), pp. 183-187, 2014.

[25].Sura Ramzi Shareef, “Breast Cancer Detection Based on Watershed Transformation”, International Journal of Computer Science, Vol. 11(1), pp.237- 245, 2014.

[26].Arnaldo Mayer and Hayit Greenspan, "An Adaptive Mean-Shift Framework for MRI Brain Segmentation", Medical Imaging, Vol. 28 (8), pp. 1238-1250, 2009.

[27].Mahalakshmi S., and Velmurugan T., “Detection of Brain Tumor by Particle Swarm Optimization using Image Segmentation”, Indian Journal of Science and Technology, Vol. 8 (22), pp. 1-8, 2015.

[28]. Venkat Narayana Rao T. and Govardhan A., "Classical and Novel Diagnosis Techniques for Early Breast Cancer Detection a Comparative Approach", International Journal of Engineering, Mathematics and Computer Sciences, Vol. 3 (11), pp. 12 -16, 2014.
[29]. Meenakshi Sundaram K., Sasikala D., and Aarthi Rani P., "A Study On Preprocessing A Mammogram Image Using Adaptive Median Filter", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 3 (3), pp. 10333-10337, 2014.
[30]. Bhausaheb Shinde and Dani A. R., "Noise Detection and Removal Filtering Techniques in Medical Images", International Journal of Engineering Research and Applications, Vol. 2 (4), pp. 311-316, 2012.
[31]. Laffont V., Durupt F., Birgen M. A., Bauduin S., and Laine A. F., "Detection of Masses in Mammography through Redundant Expansion of Scales", Proc. of the 23rd Annual International Conference EMBS (Engineering in Medicine and Biology Society), pp. 2797–800, 2001.
[32]. Meenakshi Sundaram K., Sasikala D., and Aarthi Rani P., "A Study On Preprocessing A Mammogram Image Using Adaptive Median Filter", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 3 (3), pp. 10333-10337, 2014.
[33]. Rajalakshmi M. and Subashini P., "Removal of Noise in the Chili Pepper Images Using Weighted 4- Monnected Median Filter", International Journal of Computer Science Engineering and Information Technology Research, Vol. 3(3), pp. 141- 150, 2013.
[34]. Medhat Mohamed Ahmed Abdelaal and Farouq Wael Muhamed, "Using Data Mining for Assessing Diagnosis of Breast Cancer", Proc. of International Multi conference on Computer Science, Vol. 5, pp. 11-17, 2010.

Cite This Article :

Qamar Rayees Khan, "Diagnosing Cancer using Data Mining : A Critical Review of Techniques", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 1, Issue 3, pp.94-103, November-December-2016